DOCUMENT RESUME

ED 337 462 TM 016 965

| | |
|---|---|
| TITLE | Interim Report on the Evaluation of the 1990 NAEP Trial State Assessment. April 1991. |
| INSTITUTION | National Academy of Education, Washington, D.C. |
| SPONS AGENCY | National Center for Education Statistics (ED), Washington, DC. |
| PUB DATE | 1 Apr 91 |
| NOTE | 17p. |
| PUB TYPE | Reports - Evaluative/Feasibility (142) |
| EDRS PRICE | MF01/PC01 Plus Postage. |
| DESCRIPTORS | Cost Effectiveness; *Data Collection; *Educational Assessment; Educational Policy; Elementary Secondary Education; Feasibility Studies; National Surveys; Pilot Projects; Program Evaluation; *Reliability; *Sampling; State Programs; *Validity |
| IDENTIFIERS | *National Assessment of Educational Progress; *Trial State Assessment (NAEP) |

ABSTRACT

In 1988, whether the National Assessment of Educational Progress (NAEP)--the "Nation's Report Card"--could become a report card for the states was studied. A trial assessment program was authorized to determine whether state assessments following the NAEP format could produce reliable and useful estimates of educational progress. This interim report by the National Academy of Education Panel on the Evaluation of the NAEP Trial State Assessment Project is part of a congressionally authorized evaluation of the trial assessment. The report considers the reliability and validity of the data yielded by testing a representative sample of a state's students; the utility of an indicator system such as the NAEP for guiding state policy; the effects of the state NAEP on the national NAEP; and the benefits of expanding the NAEP in light of their costs. It discusses the panel's work and its evaluation of the trials to date; the achievement levels established by the National Assessment Governing Board; the prohibition against reporting NAEP results below the state level; suggestions regarding the reauthorization of state NAEP programs; and topics for which data will be available for the October report. Interim results indicate that the sampling has gone well, without significant flaws that would threaten the integrity of the results. Recommendations are made for the release of scores from the mathematics trial assessment as scheduled, with some modifications to ensure better sampling, and an extended period for the trial program in 1994. Reauthorization by Congress is urged. (SLD)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

# NATIONAL ACADEMY OF EDUCATION

## PANEL ON THE EVALUATION OF THE NAEP TRIAL STATE ASSESSMENT PROJECT

ED 337462

THE PANEL

Robert Glaser, Chairman
Robert Linn, Co-Chairman

Gordon Ambach
*Council of Chief State School Officers*

Isabel Beck
*University of Pittsburgh*

Lloyd Bond
*University of North Carolina at Greensboro*

Ann Brown
*University of California at Berkeley*

Iris Carl
*Houston Independent School District*

David K. Cohen
*Michigan State University*

Raymon Cortines
*San Francisco School District*

Alonzo Crim
*Georgia State University*

Linda Darling-Hammond
*Columbia University*

Robert M. Groves
*U.S. Census Bureau*

Lyle Jones
*University of North Carolina at Chapel Hill*

Edward Roeber
*Michigan State Department of Education*

Albert Shanker
*American Federation of Teachers*

Lorrie Shepard
*University of Colorado*

Marshall Smith
*Stanford University*

William Winter, Former Governor
*Mississippi*

Project Director: George Bohrnstedt

TM016965

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

GEORGE W. BOHRNSTEDT

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

# April 1991 Interim Report on the Evaluation of the 1990 NAEP Trial State Assessment

**April 1, 1991**

BEST COPY AVAILABLE



Robert Glaser
Learning Research and Development Center
University of Pittsburgh
Pittsburgh, PA 15260

Robert Linn
School of Education
University of Colorado, Box 249
Boulder, CO 80309

George Bohrnstedt
American Institutes for Research, Palo
P.O. Box 1113
Palo Alto, CA 94302

**April 1991 Interim Report on the Evaluation of the 1990 NAEP Trial State Assessment**

National Academy of Education
Panel on the Evaluation of the NAEP Trial State Assessment Project
April 1, 1991

## EXECUTIVE SUMMARY

Three years ago, Congress, the Administration, the nation's Governors, the Chief State School Officers, and other education professionals determined that the time had come to see whether the National Assessment of Educational Progress (NAEP), the "Nation's Report Card", might also become a report card for the states. In the spring of 1988, Congress enacted P.L. 100-297, authorizing a NAEP Trial State Assessment (TSA) program to determine whether state assessments following the NAEP format could produce reliable and useful estimates of educational progress. As part of the authorization, Congress called for an independent evaluation of "the feasibility and validity of [state] assessments and the fairness and accuracy of the data they produce." The evaluation was to be "conducted by a nationally recognized organization (such as the National Academy of Sciences or the National Academy of Education)."

Congress viewed an evaluation of a set of trials in the states as prerequisite to the establishment of a NAEP program at the state level. Major issues to be investigated included the reliability and validity of the data yielded by testing a representative sample of a state's students; the utility of an indicator system, such as NAEP, for guiding state policy; and the effects of state NAEP, positive or negative, on national NAEP. All in all, Congress wanted to estimate the range of benefits of expanding NAEP, in light of its potential cost.

The evaluation of the TSA is being carried out under a grant from the National Center for Education Statistics to the National Academy of Education. To conduct the evaluation, the Academy appointed an independent Panel, co-chaired by Professors Robert Glaser and Robert Linn. Its first mandated report will be delivered to the Acting Commissioner of the National Center for Education Statistics in October 1991, with the purpose of providing results of the Panel's evaluation of the 1990 trial to Congress, the participating states, and the Executive Branch. The Panel has chosen to issue an interim report at this time for two reasons. First, the authorization for the TSA runs out in 1992, and it is the Panel's understanding that reauthorization hearings may begin soon. Second, the Panel believes that Congress might find its preliminary conclusions and recommendations about the 1990 trial useful, given current attention to the role of assessment in improving educational performance.

Because the first TSA results will not be released until June 1991, it is too soon to evaluate many aspects of the trial, including the various uses and impacts of the results. However, the Panel's preliminary research and deliberations provide the basis for making a set of recommendations to Congress, the states, and the Executive Branch. The Panel

believes that these recommendations can help inform the decisions Congress will soon make concerning reauthorization of state NAEP. Justification for the recommendations is presented in the attached full interim report.

Thus far, the results suggest that the 1990 trial has gone well. The Panel has not discovered any significant flaws in the sampling or administration procedures that would threaten the integrity of the results. Nor has it discovered indications that the TSA has adversely affected the national assessment. Consequently, in its role as independent evaluator of this important initiative, the Panel offers the following recommendations:

1. **On the basis of its preliminary findings from the 1990 trial, the Panel recommends the release of the state-level 1990 NAEP mathematics scale scores as scheduled.**

2. **The Panel recommends that future authorizations for state NAEP include adequate resources to sample private school students in order to increase the comparability of results from one state to another, as well as comparability to the national assessment sample.**

3. **Because of serious concerns about the validity of the achievement levels developed last fall by NAGB, the Panel recommends that NCES arrange for an independent technical review of NAGB's ongoing replication and validation studies, prior to adoption, use or reporting of achievement levels.**

4. **The use of NAEP at the school district or school level should be authorized only after careful review of policy, technical, logistical, and cost factors. The Panel plans to review such factors and recommends that the prohibition on the use of NAEP scores at the school district or school levels remain until such a review is completed.**[1]

5. **Because only two subjects at grade 4 and one subject at grade 8 will have been assessed at the conclusion of the 1992 TSA, the Panel recommends the continuance of the trial program in 1994, rather than the full establishment of a state NAEP program. Specifically for 1994, the Panel recommends trials at three grade levels -- fourth, eighth, and twelfth -- in mathematics, reading, and one additional subject, such as science.**

6. **Substantial lead-time is required for achieving national consensus on new content frameworks, and for developing assessment questions and exercises that elicit more than rote learning from students. Therefore, the Panel recommends that authority for continuation of state NAEP be made at the earliest time and that Congressional appropriations be at a level that will support appropriate assessment innovations.**

These recommendations are offered in hopes of contributing to a thorough evaluation of the promise of state NAEP. As state-level trend lines are established for achievement at various levels, in various subjects, the Panel anticipates that TSA data can become increasingly valuable to the participating states. However, the Panel wishes to register here, in addition to these recommendations, a caution against the overinterpretation from TSA results to judgments about causes or explanations of group differences in achievement. In particular, it would not be warranted from the NAEP data only to conclude that higher scores are the result of any particular differences in state policies or educational practices. As the trials move forward, it will be essential to the long-term effectiveness of this venture that those who use NAEP data exercise caution and avoid unwarranted interpretations.

1 Ambach dissents from this position; he is on record elsewhere as recommending lifting the prohibition at the school district level where the size of the enrollment enables sampling as used at the state level.

**April 1991 Interim Report on the Evaluation of the 1990 NAEP Trial State Assessment**

National Academy of Education
Panel on the Evaluation of the NAEP Trial State Assessment Project

April 1, 1991

For more than twenty years, the National Assessment of Educational Progress (NAEP) has been the best available indicator of the status of the nation's educational system. Unlike results from college admissions tests that are often used inappropriately as indicators of nationwide educational achievement, NAEP represents all students, not just a subset of college-bound high school seniors. NAEP trend data have shown that, from 1969 to the present, the average overall achievement levels in the core disciplines of reading and mathematics have been quite stable; however, achievement levels for too many students are below the levels required for their successful participation in the workforce and for the well-being of the nation. Of particular concern is the performance of 17-year-olds in science, where there has been a significant decrease in achievement over the past twenty years. Only a small proportion of students attain the basic scientific knowledge needed in this society; most fall behind early in learning science.

NAEP has been the source of some encouraging information as well. It has provided valuable insights into variations in achievement by race, ethnicity and gender. Through NAEP, policymakers learned in the 1970s and 1980s that minority students had begun to narrow the gap between their academic achievement and that of whites--though that gap remains unacceptably large.

NAEP's role as an independent indicator of educational progress is quite different from that of tests that supply information for school accountability or measure an individual student's achievement. NAEP's role is unique in that, since its first administration in 1969, it has provided the most reliable single source of information about trends in the achievement of the nation's youth. Although we may not like the discovery that levels of achievement have changed relatively little during the past twenty years and remain below those to which we aspire, NAEP will allow us to continue to monitor progress for the nation as a whole as we renew efforts for improvement and reform.

Assessment, of course, has other roles as well. The current national debate about establishing a national examination system or a national test of individual students' performances centers on using tests that would be integral to state curricula and address standards of achievement. It is critical, however, that the purposes of a national or state level indicator system be clearly distinguished from those of individual tests. NAEP was *not* designed to provide scores for individual students or schools. Indeed, such uses are precluded in the current law, which bars student identification and the reporting of results for individual schools. Ranking, comparing, or evaluating individual students, schools, or

school districts is also prohibited. In the context of current ambitions for educational change, such as a proposed national examination system and school restructuring, NAEP is best seen as an indicator that can reflect the outcomes of these changes.

## Trial State NAEP

In 1986, Secretary of Education William Bennett formed a study group to look at NAEP and to suggest ways to improve the process for assessing student achievement in the United States. Our new Secretary of Education, Lamar Alexander, served as chairman. The study group's document, The Nation's Report Card (prepared by Alexander and H. Thomas James, President Emeritus of the Spencer Foundation), recommended expanding NAEP to provide baseline and trend achievement data for the states. This recommendation was consistent with growing interest in educational progress at the state level. The report noted that primary responsibility for education in the U.S. historically has been vested in the states and argued that the value of NAEP would be enhanced if it reported state results. Participation in a voluntary state NAEP program, the report further argued, would preserve local educational autonomy and, at the same time, give states access to a core of high-quality data on performance.

The enactment of Public Law 100-297 in the spring of 1988 provided for the voluntary participation of states in NAEP on a trial basis in 1990 and 1992. In February of 1990 the first trial of the state NAEP, an assessment of mathematics achievement, was administered in more than 3,500 schools, to some 100,000 of the nation's eighth graders. In total, 37 states, the District of Columbia, and two territories participated, an indication of the wide interest in state NAEP. The results are to be released on June 6, 1991.

The second trial is scheduled for 1992, with expanded data collection to include fourth grade reading as well as fourth and eighth grade mathematics. The continuation of the trials through 1994, however, is contingent on Congressional action.

## The Panel's First Year of Activity and the Reasons for this Report

Public Law 100-297 also mandated that an independent evaluation be conducted to assess the feasibility and validity of the Trial State Assessments (TSA). In October 1989, the National Center for Education Statistics (NCES) commissioned the National Academy of Education (NAE) to conduct this evaluation. The NAE assembled a panel of experts in a broad range of technical and policy fields in education and arranged for technical and staff support from the American Institutes for Research (AIR). The Academy's panel held three meetings in its first year. During this time, the Panel focused on the data and information it would require, and the design of studies needed to conduct an effective evaluation. The

Panel has made major decisions on a first-phase agenda for obtaining information about the 1990 trial, its impact, and questions related to the future value and validity of the TSA.

The key questions that guide the Panel's work are those Congress, in 1988, anticipated would be crucial to evaluating the TSA: (1) How well was the assessment implemented from a technical perspective? (2) How valid and accurate is the assessment? Has it yielded valid and reliable data at the state level? (3) How useful are the results and reports generated from the assessment? To answer these questions, the Panel has commissioned a set of studies and papers to address prominent aspects of the TSA design and implementation. The results of these inquiries will clarify the appropriate role of state indicators. They will also reveal how state NAEP might help monitor progress toward national educational goals.

Because authorization hearings beyond 1992 may begin soon, the Panel has chosen to issue this interim report, prior to releasing its first mandated report in October. Although it is still early in the evaluation, the Panel's findings and discussions to-date have direct bearing on issues that may be considered in connection with future authorizations. These findings can also inform national deliberations about the use and effects of educational assessments, particularly those relating to the work of two groups: the National Education Goals Panel and the President's Educational Policy Advisory Committee.

The balance of this interim report includes the following sections: A review of the Panel's work, along with its evaluation of the trials to this point in time; discussion of the achievement levels established by NAGB; discussion of the prohibition against reporting NAEP results below the state level; suggestions regarding the reauthorization of state NAEP; and a short overview of topics for which data will be available for the October report.

## The Evaluation of the Trial To-Date

At present, the Panel is prepared to offer preliminary observations about the 1990 TSA based on data and deliberations in four areas: (1) sampling, (2) excluded student populations, (3) administration, and (4) inferences that can and cannot be made from the 1990 Trial State Assessment. A more complete and detailed report on work in these areas will be presented for Congressional, state, and Executive Branch consideration in October.

### Sampling

The preliminary analysis of the sampling design and its execution has focused on the reports on sampling in the February 1990 trial that were available as of January 1991. From this analysis the Panel has concluded that the sampling was competently performed. A common difficulty encountered in programs such as NAEP is nonparticipation; some

schools refuse to participate and some students either refuse or are absent. The magnitude of nonparticipation in the TSA as a whole was reasonably small, with about 6% of the schools declining to join the project and about 6% of the students in the participating schools not taking part. These rates varied from state to state, however, and in two states the rate of students not participating was between 10% and 20%. Statistical adjustments, known as "nonresponse adjustments," are being used to compensate for the missing data; the adjustments appear to be appropriate and reasonable. In sum, school and student participation in the 1990 sample produced a generally favorable picture for state NAEP.

In considering issues of sample design, the Panel has been alert to any indication that state NAEP might have a negative impact on participation in national NAEP. Thus far we have found no cause for concern. While it is true that state NAEP did increase the burden on small states to provide a sufficient number of schools to meet the requirements of both the national and state sampling frameworks, few of the 37 states that participated in the first trial found it to be a problem. Furthermore, there was no indication in the administration of the 1990 trial of interference with the administration of national NAEP. In sum, the Panel can report that, thus far, the 1990 TSA has had no discernible negative impact on the 1990 national NAEP.

**Excluded Student Populations**

The design of the TSA allowed for the exclusions of three groups of students: students enrolled in private schools, students with limited English proficiency (LEP), and special education students with individualized education plans (IEP). Together, these groups make up about one-sixth of the eighth grade nationally. Differences among the states in the proportions of students in these groups could have important effects on state-by-state and state-to-national NAEP comparisons. For example, in national NAEP, which tests both private and public school students, private school students tend to perform better than public school students. Because comparisons of states' performances on NAEP inevitably will be made, the exclusion of private school students in the TSA is cause for concern; states' performances could change substantially with inclusion of the private school students.

**Private School Students.** Private school students typically made up the largest excluded group in each state. Nationally, about 12% of the eighth grade students are enrolled in private schools, but this percentage varies widely across states: In seven states fewer than 5% of the eighth grade students are in private schools, and in seven others more than 18% are in private schools. Wyoming and Utah each enroll only about 2% in private schools, whereas Hawaii and the District of Columbia each enroll about 20%.

At this time the Panel does not know how much the inclusion of private school students would affect the rankings of states. The magnitude of the effect depends on how many students in a state are enrolled in private schools and on the size of the differences

between public and private school students' performances on the NAEP items. For the October report, the Panel is conducting analyses to examine how the states' results might change as a function of the exclusion of private school students.

The Panel believes that state NAEP data would better reflect educational achievement and make state results more readily comparable if, in addition to results for public school students, results for all students (in public and private schools) were produced.

**Limited English Proficiency (LEP) and Individualized Education Plan (IEP) Exclusions.** The exclusion criteria for LEP and IEP students were part of the sampling design and were implemented locally, but the local administrators were told to include doubtful cases in the assessment. Overall, about 1% of the students were excluded for reasons of limited English proficiency and about 4% because they had individualized education plans. But the percentages across states varied, with LEP exclusions ranging from near 0% in West Virginia and about 2% in New Jersey, New York, Rhode Island, and Texas, to 5% in California. IEP exclusions ranged from about 2% in Montana to 8% in Arkansas.

The Panel is in the process of examining how consistently the exclusion rule was implemented in the first trial state assessment, but findings will not be available until the Panel's October report.

## Administration

As part of the study on administration, Panel staff directly observed training sessions for test administrators and assessment sessions. In addition, they conducted independent analyses of the TSA data and Quality Control Monitoring Data collected by the NAEP contractor. Finally, they conducted a survey of State Testing Directors as an independent appraisal of the 1990 TSA administration. Their initial findings focus on issues of local conditions for implementing the TSA, the impact of the first trial on the 1990 national NAEP, and state testing directors' responses to the first trial.

**Local Conditions.** Unlike national NAEP, the TSA employed local school staff to administer the test sessions, rather than staff employed by the contractor. Uniformity of assessment conditions is a prerequisite for the legitimate comparison of a state's results, both with the national composite result and with the results of other states. The administration of TSA by local staff had the potential to threaten the comparability of state results with national results. Although the local test administrators underwent careful training, there was the possibility that students might respond differently when tested by them, and that this would distort the results of the TSA. Therefore, the NAEP contractor had a monitor present in a random half of the test sessions to assure that the local administrators proceeded according to their training.

The critical finding was that student performance in monitored sessions did not significantly differ from the performance in unmonitored sessions, suggesting that local administrators were largely successful in implementing uniform testing conditions that did not advantage or disadvantage students. Quality control monitors looked for variations in every aspect of the testing session, including timing, reading the script, and handling student questions. Analysis of the reports indicates that deviations from uniform procedures were infrequent and were unlikely to have systematically influenced state results.

**The Trial and National NAEP.** Because of possible differences in testing conditions, a second issue was whether students assessed in the TSA might obtain scores that, on average, differ from those obtained by students in national NAEP. Since the sampling frame for the TSA differed from that of national NAEP, the contractor constructed a "matched" subset of the students from national NAEP to enable valid comparisons. Compared to the matched subset of national NAEP students, students in the TSA obtained slightly, but reliably, higher scores. While the design of the study prohibits a definitive explanation for the difference, the Panel is exploring the possibility that students participating in state NAEP may have been more motivated to do well than those in national NAEP. The Panel will report further on this issue in its October report. In addition, the Panel will closely monitor the 1992 trials to see if this potentially important finding is replicated.

**TSA Planning and Policy from the State Testing Directors' Perspective.** State testing directors are knowledgeable observers and important stakeholders in the assessment process. To monitor their responses to the first trial, the Panel staff conducted an independent survey of the directors as part of the study on administration. Most reported that the assessment went well and that the data from the TSA would be of value to their states. However, some noted that they felt excluded from important policy decisions in the establishment and implementation of the TSA. The Panel applauds the efforts of NAGB and NCES in their stated intention to use CCSSO's Education Information Advisory Committee as a vehicle for providing state testing directors with greater policy input. However, because the Panel recognizes the need for close cooperation among NCES, the contractor, and the states participating in the TSA, the Panel proposes that the governance and administrative structures of NAEP strengthen the mechanisms for securing input from state testing directors into the state NAEP policy and assessment development process.

Based on the preliminary results of its studies and its ongoing deliberations, the Panel believes that the 1990 TSA has proceeded well. Thus far, the studies have identified no signs that the experiment is flawed, that major redirection is necessary, or that the TSA should be terminated. **On the basis of its preliminary findings from the 1990 trial, the Panel recommends the release of the 1990 NAEP mathematics scale scores as scheduled.**

Despite this generally favorable observation, the Panel is concerned that the exclusion of private school students from the TSA ultimately will diminish the utility of the

trials and future administrations of the state NAEP. Given both the significant variation from state-to-state in the size of this group, and its inclusion in the national sample, issues of comparability become much more complex than need be when private school students are excluded from the sample. **The Panel recommends that future authorizations for state NAEP include adequate resources to sample private school students in order to increase the comparability of results from one state to another, as well as comparability to the national assessment sample.**

**Inferences That Can and Cannot Be Drawn from the Trials**

Congress should be aware of the kinds of inferences that can be usefully drawn from the TSA, given the design of the assessments. As state-level trend lines are established for achievement at various levels, in various subjects, state NAEP data will become increasingly valuable to the participating states. They will provide governors, legislators, and state school officials with the ability to monitor educational progress using information of unparalleled richness. These trend lines will enable comparisons with similar states, the nation, and other countries as the basis for much-needed educational innovation. The two data points for eighth grade mathematics provided by the 1990 and 1992 TSAs will provide valuable preliminary trend information to those states that participate in both trials. However, the real value will come with the accumulation of additional data points across time.

The ability to compare similar states will prove useful in the consideration of policy issues. It should be emphasized, however, that the results will not support causal inferences about what produced differences in achievement. In particular, it would not be safe to conclude that higher scores are the result of any particular differences in state policies or educational practices.

At this juncture, it is important to remember that the 1990 Trial State Assessment is limited in scope: it embraces only one subject at one grade level, eighth grade mathematics. With the inclusion of fourth grade reading, fourth grade mathematics, and eighth grade mathematics in the 1992 trial, policymakers and the public will have a valuable, yet narrow, window on learning outcomes across the two grade levels and curriculum areas. The Panel cautions against overgeneralization from these trials to questions of schools' and teachers' performances or group differences in achievement.

The results will see their best use in the establishment of trends in achievement within a state, over time, and in the drawing of comparisons between states with similar populations, and between a state and the nation. However, comparisons of states' rankings inevitably will be made. While states can be ranked with respect to mean levels of achievement, interpretations of state-to-state differences must be made with great caution. Three issues must be addressed. (1) It must be determined whether the differences between the rankings are large enough to be considered reliable. (2) The relevance of a

state's ranking to judgments about its educational quality will depend upon the match between the content tested by NAEP and the state's curriculum framework as implemented. Some state frameworks are closer than others to the content of NAEP. (3) Differences in states' performances may be due to differences in demographics. The Panel has studies in place to examine all three of these issues.

## Issues Currently Under Discussion and Debate

Since the Panel received its mandate for the evaluation of the 1990 TSA from NCES, two important policy issues relevant to state NAEP have become prominent. The first is the proposal for and the development of a set of achievement levels or standards, using the 1990 NAEP mathematics items. The second is a recommendation by the National Assessment Governing Board for lifting the current prohibition against the reporting of NAEP results below the state level. Given the importance of both these issues for state NAEP, the Panel has agreed to address them in this report.

### "Standards" or Achievement Levels

The legislation that authorized State NAEP (P.L. 100-297) also assigned to NAGB the task of developing appropriate achievement goals for each age, grade, and subject area in NAEP. The unveiling of a set of six educational goals by the White House and the Governors in 1989 heightened interest in educational standards, and set the stage for NAGB to develop a set of achievement levels that *could* be used to measure progress toward the national goals. Last August, NAGB engaged in an exercise to define basic, proficient, and advanced achievement levels in fourth, eighth, and twelfth grades, using the 1990 NAEP mathematics items. The Panel applauds this attempt to make scores more interpretable, but cautions that it must be viewed as an intricate process involving judgment, definition, and, ultimately, issues of reliability and validity.

As valuable as achievement levels might be for the states in monitoring their progress toward meeting some of the national educational goals, the results of the process of setting the achievement levels should meet the scrutiny of experts and be credible to the public. The Panel concurs with NCES's Technical Review Panel and CCSSO that the current achievement levels, obtained before January 1991, are flawed. As a result, the Panel's Chairman and Co-chairman have written to Richard Boyd, Chair of NAGB, urging that the achievement levels be used only if corrected. NAGB is in the process of conducting a replication and validation study in four regions of the country. The Panel commends NAGB's efforts to secure validation of the achievement levels, since the data collected for that purpose should be adequate for evaluating the current levels, or if necessary, modifying or discarding them. Since the Panel believes that the use of inadequately developed achievement levels could have a corrosive effect on state participation in the future, as well as on the credibility of NAEP more generally, the Panel will monitor the validation studies.

**Because of serious concerns about the validity of the achievement levels developed last fall by NAGB, the Panel recommends that NCES arrange for an independent technical review of NAGB's ongoing replication and validation studies, prior to adoption, use or reporting of achievement levels.**

### Reporting TSA Data Below the State Level

NAGB recently has recommended to Congress that the current prohibition against reporting NAEP data below the state level be lifted to allow reporting at the school district or school. NAGB would continue the prohibition on reporting individual student scores. The Panel supports NAGB's recommendation to continue the prohibition against reporting data at the student level. But the Panel also believes that expansion of NAEP to provide results at the individual school building level or for other than large school districts could lead to the loss of NAEP as an independent and uncorrupted indicator of educational progress. NAEP's historic role as an auditor that stands apart from the training and testing of individual students can too easily be compromised by its use at the school and student level.

The extension of NAEP to the district level raises a somewhat different set of issues. The reporting of data for at least some of the largest districts may be as warranted as reporting data for some of the smallest states. Indeed, because of its special status, the District of Columbia did participate in the 1990 TSA. Prior to lifting the prohibition, however, the Panel believes that the technical, policy, and cost implications, as well as the implications for future test design and administration, need careful study and consideration. The Panel plans to commission a study on the implications of reporting NAEP data below the state level and will present the results and conclusions in a future report. **The use of NAEP at the school district or school level should be authorized only after careful review of policy, technical, logistical, and cost factors. The Panel plans to review such factors and recommends that the prohibition on the use of NAEP scores at the school district or school levels remain until such a review is completed.**[1]

### Panel Perspectives on Key Issues in Reauthorization

The Panel recognizes the great value of maintaining continuity of state NAEP, especially in light of the general technical success of the 1990 trial. But as Congress considers reauthorization of 1994 NAEP, the Panel suggests a number of important issues to consider.

**Planning the 1994 Trial State Assessment**

The Panel recommends that 1994 NAEP, when reauthorized, should include additional state trials since, with the conclusion of the 1992 trial, only two subjects, mathematics and reading, will have been evaluated at two grade levels. In 1994, national NAEP will assess mathematics, reading, science, and history and geography combined. Authorizing state trials for one subject (e.g., science) in addition to reading and mathematics and for an additional grade level (twelve in addition to four and eight) prior to moving to a fully implemented state NAEP would be informative. By 1994, trends for fourth grade mathematics and reading would be available in addition to the trends for eighth grade mathematics for 1990 and 1992, thereby allowing for a more complete evaluation of the uses of and the interest in such trend data by the participating states. In addition, such an expansion would provide data to help evaluate the feasibility, impact, and cost of a fully implemented state NAEP.

The Panel suggests the addition of a twelfth grade trial in 1994. Of central importance to the Panel is the fact that results from the trials at the fourth and eighth grade levels cannot be assumed to generalize to the twelfth grade. The motivation of twelfth graders to participate and perform well may be very different from that of students in the lower grades. Moreover, state level results for twelfth grade students may be of particular interest and use to the states. There is great concern about workforce preparedness on the part of private industry, the Administration, Congress, and the states.

Finally, preliminary evaluation results suggest that the 1990 trial is going well. However, before the Panel can reach a final conclusion regarding the success of the trials, it must complete its evaluation of the 1990 and the 1992 trials. There is much useful information to be gained from continuing the trial program to inform the fuller development and implementation of state NAEP in the longer term.

**Because only two subjects at grade 4 and one subject at grade 8 will have been assessed at the conclusion of the 1992 TSA, the Panel recommends the continuance of the trial program in 1994, rather than the full establishment of a state NAEP program. Specifically for 1994, the Panel recommends trials at three grade levels -- fourth, eighth, and twelfth -- in mathematics, reading, and one additional subject, such as science.**

**Assuring the Quality of State NAEP**

With Congress' requirement in 1988 that a national consensus process be carried out when updating test content frameworks, NAEP has reaffirmed its status as an innovator. The 1992 reading assessment reflects the current emphasis on performance-based assessment, and the 1994 science assessment seems likely to pursue the same progressive route. The Panel believes that NAEP should exemplify and promote current

innovations in assessment technology on a stage-by-stage basis. To provide for trend data, provision must be made for assessments to include items that maintain links to past assessments and, at the same time, build links to the future. The consensual development and updating of content frameworks are essential to securing innovation and planning this balance.

New assessment technologies and innovations carry with them increased costs and require considerable time to develop. The Panel is also aware of the massive amount of work that must be completed in relatively short periods in the implementation of NAEP. For example, CCSSO had less than four months in 1989 to create the reading framework and must, in seven months, create the new science framework. Working within such schedules, while incorporating high-quality innovations in assessment technology, is nearly impossible.

**Substantial lead-time is required for achieving national consensus on new content frameworks, and for developing assessment questions and exercises that elicit more than rote learning from students. Therefore, the Panel recommends that authority for continuation of state NAEP be made at the earliest time and that Congressional appropriations be at a level that will support appropriate assessment innovations.**

## The Panel's October Report

The Panel's mandated report in October will expand on the topics addressed in this interim report, and will focus as well on: the presentation and impact of the results of the 1990 Trial, the content validity of the items, and the policy context of goals for achievement in which the TSA is embedded.

The results of the 1990 Trial State NAEP will be released on June 6, 1991. The Panel is interested in the clarity, interpretability, and usefulness of different formats for reporting results to the states. It will also investigate any moves toward curricular or instructional changes in states' mathematics programs. Finally, it will examine the degree to which the reports are fair--that is, the degree to which the rankings of states vary as a function of different types of test content (e.g., algebra versus geometry), or as a consequence of adopting alternative methods for producing an overall score. The Panel will also examine the relation between state assessment results and the racial, ethnic, and gender composition of the states.

When Congress authorized NAEP in P.L. 98-511, it required that the curriculum frameworks be developed through a national consensus process, providing for the participation of teachers, curriculum specialists, school administrators, parents, and members of the general public. In October, the Panel will report on the adequacy of this

consensus process for the 1990 mathematics assessment and the 1992 reading assessment. The report will describe the constituencies represented and the nature of the advice sought. It will also evaluate how this advice and input affected the design of the frameworks, and the extent to which the frameworks represent a consensus among professionals in the fields of mathematics and reading education. Of particular interest for TSA is the degree to which the consensual process represents a national perspective that includes the current goals and objectives of state and local school districts.

There is considerable interest in using the results from the achievement levels for inclusion in the state "report cards" that the National Goals Panel will release this September. In addition, discussion continues about whether there should be a national examination, and if so, what role NAEP and state NAEP should play if a national examination is established. The Panel continues to monitor the policy context in which the 1990 Trial is occurring, and will report more fully on that context in its October report.

The Panel hopes that this interim report regarding the 1990 trial in mathematics and TSA reauthorization will prove useful to Congress as it deliberates about the future of state NAEP. The recommendations endorsed here will allow thorough evaluation of its promise as a valuable indicator of states' educational achievement and will strengthen the possible full extension of NAEP to the states. Over the shorter and longer terms, state NAEP may serve as a vital measure of progress toward the achievement of the educational goals that are a priority for the states and the nation.

[1] Ambach dissents from this position; he is on record elsewhere as recommending lifting the prohibition at the school district level where the size of the enrollment enables sampling as used at the state level.